7th Framework Programme
ENV.2010.4.1.2-2
Integrating New Data Visualisation Approaches of Earth Systems into GEOSS Development

Project Nr: 265178

QUALity aware VISualisation for the Global Earth Observation system of systems

Deliverable D4.1
Graphical search interface report

Version 1.0

Due date of deliverable: 31/07/2013
Actual submission date: 10/11/2013
<table>
<thead>
<tr>
<th>Document control page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Creator</strong></td>
</tr>
<tr>
<td><strong>Editor</strong></td>
</tr>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
</tr>
<tr>
<td><strong>Contributors</strong></td>
</tr>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td><strong>Format</strong></td>
</tr>
<tr>
<td><strong>Language</strong></td>
</tr>
<tr>
<td><strong>Creation date</strong></td>
</tr>
<tr>
<td><strong>Version number</strong></td>
</tr>
<tr>
<td><strong>Version date</strong></td>
</tr>
<tr>
<td><strong>Last modified by</strong></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
</tr>
<tr>
<td><strong>Dissemination level</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Nature</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Review status</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Action requested</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Requested deadline</strong></td>
</tr>
</tbody>
</table>
Revision history

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Modified by</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>09/06/2013</td>
<td>CREAF_JM</td>
<td>Initial Draft</td>
</tr>
<tr>
<td></td>
<td>03/07/2013</td>
<td>CNR_LB</td>
<td>Introduction and general revision</td>
</tr>
<tr>
<td></td>
<td>11/07/2013</td>
<td>CREAF_JM</td>
<td>Section 3</td>
</tr>
<tr>
<td></td>
<td>01/08/2013</td>
<td>CNR_RR</td>
<td>Section 2</td>
</tr>
<tr>
<td></td>
<td>02/10/2013</td>
<td>52N_DN</td>
<td>Section 3</td>
</tr>
<tr>
<td></td>
<td>03/10/2013</td>
<td>CNR_RR</td>
<td>Integration of contributions</td>
</tr>
<tr>
<td></td>
<td>09/10/2013</td>
<td>CNR_LB</td>
<td>Revision and consolidation</td>
</tr>
<tr>
<td>28/10/2013</td>
<td>CREAF_EP, CREAF_JM</td>
<td>Section 4 and recommendations to Section 5</td>
<td></td>
</tr>
<tr>
<td>1.0-Final Draft</td>
<td>28/10/2013</td>
<td>CNR_RR</td>
<td>Integration of contributions</td>
</tr>
<tr>
<td>4/11/2013</td>
<td>52N_DN</td>
<td></td>
<td>Revision of Section 3 and recommendations in Section 5</td>
</tr>
<tr>
<td>8/11/2013</td>
<td>CNR_RR</td>
<td></td>
<td>Integration of comments and finalization</td>
</tr>
<tr>
<td>10/11/2013</td>
<td>CREAF_EP, CREAF_JM</td>
<td>Additions to Section 4. Final version submitted</td>
<td></td>
</tr>
</tbody>
</table>

Institution | Contributors
---|-------------------
CNR_LB | CNR: Lorenzo Bigagli
CNR_RR | CNR: Roberto Roncella
52N_DN | 52N: Daniel Nüst
CREAF_EP | CREAF: Ester Prat
CREAF_JM | CREAF: Joan Masó

Copyright © 2013, GeoViQua Consortium

The GeoViQua Consortium grants third parties the right to use and distribute all or parts of this document, provided that the GeoViQua project and the document are properly referenced.

THIS DOCUMENT IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS DOCUMENT, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
# Table of Contents

List of Figures .......................................................................................................................................................... 6

List of Tables ............................................................................................................................................................ 7

Executive Summary ...................................................................................................................................................... 8

1. Introduction .......................................................................................................................................................... 9
   1.1 Scope ............................................................................................................................................................... 9
   1.1.1 Relation to GeoViQua tasks .................................................................................................................... 9
   1.2 GeoViQua enhanced geo-search components architecture ............................................................. 9
   1.3 Conventions .................................................................................................................................................. 10
   1.3.1 Acronyms and abbreviations ............................................................................................................... 10

2. GeoViQua DAB-Q Demo Portal ......................................................................................................................... 12
   2.1 GI-portal before the project .................................................................................................................... 12
   2.2 Enhancements to GI-portal ..................................................................................................................... 16
   2.2.1 Improvements to the query constraints ........................................................................................... 16
   2.2.1.1 Support to query the PQM ....................................................................................................... 18
   2.2.1.2 Support to query the UQM ..................................................................................................... 20
   2.2.2 Improvements to quality-aware services ........................................................................................ 20
   2.2.3 Improvements for feedback support .................................................................................................. 21

3. GeoViQua GEO Portal Mirror .......................................................................................................................... 24
   3.1 GEO Portal before GeoViQua .................................................................................................................. 24
   3.2 GEO Portal improvements ....................................................................................................................... 26
   3.2.1 Modifications in the Search Results / Dataset Listing page .......................................................... 27
   3.2.1.1 Visual Metadata Summary ..................................................................................................... 27
   3.2.1.2 Result Item Actions ................................................................................................................. 27
   3.2.2 Modifications in the Single Resource / Dataset View page .......................................................... 30
   3.2.3 Modifications in the Refine This Search page ................................................................................ 31
   3.2.4 Extensions of Web Map Browsing ................................................................................................... 32

4. Search assisted by statistical plots ...................................................................................................................... 34
   4.1 Including search by statistical plots in the GEO Portal ................................................................. 34
   4.1.1 Spatial Data Infrastructure of Catalonia catalogue portal ............................................................ 34
   4.1.2 The INSPIRE Geoportal ................................................................................................................. 37
   4.2 Proposals on the use of statistical plots ................................................................................................. 40
   4.2.1 Parallel coordinates plot ................................................................................................................... 41
   4.2.2 Scatter plots ........................................................................................................................................ 42
   4.2.3 Star plots .......................................................................................................................................... 42
   4.2.4 Chernoff faces .................................................................................................................................. 43
   4.3 Factors to consider when implementing statistical plots ...................................................................... 43
   4.4 The GEO label as a statistical plot for metadata completeness ...................................................... 44
   4.5 Example of implementation in the GEO Portal Mirror GUI ........................................................... 45

5. Conclusions .......................................................................................................................................................... 50
   5.1 Recommendations for GCI integration ................................................................................................. 50
   5.1.1 Recommendations about the inclusion of GeoViQua broker ....................................................... 50
   5.1.2 Recommendations about inclusion of statistical plots .................................................................. 50
   5.1.3 Recommendations about inclusion of the user feedback ........................................................... 50
5.1.4 Recommendation about inclusion of GEO label .................................................................50

5.2 Future improvements .............................................................................................................51
  5.2.1 Query by label in the DAB-Q Demo Portal .................................................................51
  5.2.2 Customizable statistical plots ......................................................................................53

6. References ............................................................................................................................54
List of Figures

Figure 1 – Simplified Components Diagram ................................................................. 10
Figure 2 – GI-portal before the start of GeoViQua .......................................................... 12
Figure 3 – GEOSS DAB Source selection .................................................................... 13
Figure 4 – Query status ................................................................................................. 14
Figure 5 – Query results ............................................................................................... 14
Figure 6 – Resource metadata ..................................................................................... 15
Figure 7 – DAB-Q Demo Portal .................................................................................... 16
Figure 8 – Quality and feedback query options .............................................................. 17
Figure 9 – Quality-enabled service instances ............................................................... 21
Figure 10 – Example query results ................................................................................ 22
Figure 11 – User Feedback editor window ................................................................. 23
Figure 12 – Current GEO Portal entry page ............................................................... 24
Figure 13 – Search results page ..................................................................................... 25
Figure 14 – “Refine this search” page .......................................................................... 25
Figure 15 – Single result page ....................................................................................... 26
Figure 16 – Current GEO Portal map viewer ............................................................... 26
Figure 17 – Preliminary search results page (GEO Label option) .................................. 27
Figure 18 – Extended search result page .................................................................... 28
Figure 19 – Metadata comparison ............................................................................... 29
Figure 20 – Single resource page with new items for GEO label, user feedback and quality-enabled viewers. ............................................................... 30
Figure 21 – Refine search page .......................................................................................... 32
Figure 22 – Greenland map analysis client in the Geoportal .......................................... 33
Figure 23 – a) Top of the results page with only the few first items visible. b) Bottom of the results page where a great amount of pages containing lists of results available are displayed. ............................................................... 35
Figure 24 – Statistical distribution of different category values .................................... 36
Figure 25 – Spatial distribution for different datasets .................................................... 37
Figure 26 – Home page of the INSPIRE Geoportal ......................................................... 37
Figure 27 – Results of the application of a geographical filter by drawing a bounding box. ............................................................... 38
Figure 28 – View of the pop-up menu in the text search box showing the multilingual thesaurus ... 39
Figure 29 – Advanced search provides values for each category and number of occurences. ............................................................... 39
Figure 30 – Sorting options .......................................................................................... 40
Figure 31 – Bar charts as filtering tool .......................................................................... 41
Figure 32 – Parallel coordinates plot example where high values mean better criteria. ............................................................................................... 42
Figure 33 – Scatter plot example where points at the top right corner mean better criteria ............................................................................................... 42
Figure 34 – Star plot example where external values (more area) mean better criteria. ............................................................................................... 43
Figure 35 – Chernoff faces example where smiling faces means better criteria. ............................................................................................... 43
Figure 36 – Example of the possible appearances of the sections in the GEO label. Each section will have one of the three independent appearances depending on the availability of the information for this case. ............................................................... 45
Figure 40 – Search interface with disabled GEO label filter ........................................... 51
Figure 41 – DAB-Q Demo Portal with enabled GEO label filter ..................................... 52
Figure 42 – DAB-Q Demo Portal with selected GEO label facets .................................. 53
List of Tables

Table 1 – Classification and definition of quality elements by ISO 19115 ...........................................18
Table 2 – Producer Quality Model (PQM) queryables ........................................................................19
Table 3 – User Quality Model (UQM) queryables ..............................................................................20
Executive Summary

This document describes the GeoViQua Graphical User Interface (GUI) for improved quality-enabled discovery of geospatial resources in the context of GEOSS.

This deliverable presents two user interface implementations: the GeoViQua DAB-Q Demo Portal interface, based on the GI-portal client of the EuroGEOSS Broker; and the GeoViQua GEO Portal Mirror interface, based on the GEO Portal.

Besides, it suggests how to implement a search with the support of statistical plots based on the available data in the DAB.

As a contribution to GEOSS, this deliverable includes recommendations on improvements to the GEOSS Common Infrastructure and to the official GEO Portal, to implement the functionalities of the GeoViQua GUI.
1. Introduction

1.1 Scope

The present document is the Deliverable 4.1 “Graphical search interface report” of the GeoViQua Project. [GVQ-DOW 2010]

It describes the GeoViQua Graphical User Interface (GUI) for improved quality-enabled discovery of geospatial resources in the context of GEOSS.

The proposed solution has been experimented in a twofold manner:
- As an extension of the GI-portal\(^1\) interface, developed in the framework of the EuroGEOSS project;
- As an extension of the official GEOSS GEO Portal\(^2\).

Indications and guidance have been gathered from the users and the project partners by a number of means, including project deliverables, formal and informal meetings and other communications.

The GeoViQua GUI is presented as a contribution to GEOSS, hence this document includes recommendations on improvements to the GEOSS Common Infrastructure and to the official GEO Portal, to implement the functionalities of the GeoViQua GUI.

This document is mainly targeted at a technical audience and is concerned with ICT aspects only; all the other concerns (e.g. quality of data and scientific methods, legal aspects) are out of scope.

This document is structured in four main sections. This Section 1 provides an introduction to the project Work Package 4 (WP4) and its background, and establishes the terminology and notation used throughout the rest of the document. Section 2 describes the GeoViQua DAB-Q Demo Portal GUI, based on GI-portal. Section 3 describes the GeoViQua GEO Portal Mirror GUI, based on GEO Portal. Section 4 makes suggestion on how to implement a search assisted by statistical plots, based on the available data in the DAB and on a protocol to get these statistical overviews. Conclusions and recommendations on improvements to the GEOSS Common Infrastructure and to the official GEO Portal are presented in Section 5.

1.1.1 Relation to GeoViQua tasks

This deliverable is directly related to Task 4.2 “Quality-aware catalogue client” of WP4 “Enhanced geo-search components”.

It relates to: WP3 “Data quality elicitation mechanisms”, particularly to Task 3.5 “User feedback and quality assessment of data sets in GEOSS”, which defines the GeoViQua data model; and to WP6 “Delivery of solutions to end users”, which integrates everything.

1.2 GeoViQua enhanced geo-search components architecture

Figure 1 shows the simplified component diagram and description of the architecture.

---

\(^1\) http://www.eurogeoss-broker.eu/
\(^2\) http://www.geoportal.org/web/guest/geo_home
Task 4.2 description
This task will develop a client component to query the catalogue service developed in the Task 4.1.
- The client will allow to express constraints based on typical geo-information parameters (e.g. the W4: Where, When, What, Who), GEO relevant parameters (e.g. scale/detail, SBAs and categories), examples of specific parameters for GEOSS SBAs applications (e.g. taxonomies for biodiversity), quality information/metadata. Quality parameters will be expressed through either metadata constraints, or through relationships between dataset layers and quality layers. The client will be web-based client to be included in Web applications and portals. The client will also implement extended functionalities to get feedback information about query distribution, query mediation, results translation, ensemble and presentation. [CNR, 52N]
- Data mining will be improved form the user prospective by providing the user some graphical (e.g. bar charts) representation of the datasets available based on different metadata parameters. These graphics give a general picture of the dataset collection to the user before he starts to make requests. Furthermore, users will be able to dynamically filter the available datasets (e.g. by selecting a bar in the chart) and other graphics will be updated with the subset filtered. The process can be repeated by the user reducing the collection to a subset with a size that makes easier to screen it by a classical textual request. [CREAF]

1.3 Conventions

1.3.1 Acronyms and abbreviations
CSW – Catalogue Service for the Web
DAB – Discovery and Access Broker
FP7 – Seventh EU Framework Programme for Research and Technological Development
GCI – GEOSS Common Infrastructure
GEO – Group on Earth Observation
GEOSS – Global Earth Observation System of Systems
GML – Geography Markup Language
GUI – Graphical User Interface
GVQ – GeoViQua
ICT – Information and Communication Technology
IDEC – Spatial Data Infrastructure of Catalonia
ISO – International Organization for Standardization
KML – Keyhole Markup Language
KVP – Key-value pair
OGC – Open Geospatial Consortium
OWS – OGC Web Service
SOS – Sensor Observation Service
UML – Unified Modeling Language
WAF – Web Accessible Folder
WCS – Web Coverage Service
WFS – Web Feature Service
WMS – Web Map Service
WP – Work Package
WPS – Web Processing Service
2. GeoViQua DAB-Q Demo Portal

The GeoViQua Catalogue is an instance of the GeoViQua quality-enhanced catalogue broker (DAB-Q), that will be documented in the deliverable D4.2. The Demo Portal is a demonstration client that we have developed to showcase its features, extending the test GI-portal\(^3\) interface realized in the framework of the EuroGEOSS project.

The EuroGEOSS project adopted a brokering approach to implement multi-disciplinary interoperability and lower entry barriers for both Users and Data Providers. According to such approach, Users and Data Providers are not asked to implement any specific interoperability technology, but can continue using their tools and publishing their resources according to their standards. A layer of higher-level services provides the capabilities to harmonize and integrate such resources. The components implementing such services constitute the EuroGEOSS Brokering Framework, or EuroGEOSS Broker. The EuroGEOSS Broker has been incorporated into the GCI and renamed to GEO Discovery and Access Broker (GEO DAB).

2.1 GI-portal before the project

GI-portal is a web-client interface for querying the many heterogeneous catalogues connected to the GEO DAB. Figure 2 shows the interface of GI-portal before the start of GeoViQua (please note that the currently deployed instance of GI-portal may have evolved since then).

Figure 2 – GI-portal before the start of GeoViQua

The GEO DAB can access a multiplicity of catalogue services, as well as inventory and access services, through one or more web service interfaces, such as OpenSearch and the OGC Catalog service for the Web (CSW), to discover (and possibly access) heterogeneous geospatial resources.

\(^3\) http://www.eurogeoss-broker.eu/
Currently, the GEO DAB supports several OGC Web Services (WCS, WMS, WFS, WPS), THREDDS Data Server, SeaDataNet Common Data Index, the GEOSS Clearinghouse (based on GeoNetwork), and others. By default, the GEO DAB distributes the queries to the whole GCI. If desired, individual data sources can be selectively excluded or included (see Figure 3).

Queries can be formulated by composing the desired filters corresponding to basic constraints (e.g. Where, What, When, Who) by means of widgets in the right part of the GI-portal GUI. Once the desired constraints are selected (they are all optional), users can start the query by clicking the “Start search” button at the bottom right.

A query status window will display feedback from all the sources that are involved in the query (Figure 4).
The query results will be displayed on the map and listed in the bottom frame (Figure 5).
Each row contains information on a specific matching resource, as well as GUI controls to perform further actions on it. A short description of the query results columns follows:

- **Access/Use Constraint**: information on constraints that apply to access and/or use the resource; for example if access to the resource is restricted, or follows the GEOSS data core sharing principles.
- **Helper buttons**: allow launching helper applications to perform further elaborations on the selected resource. For example, the GI-axe access broker supports the download of the selected resources according to a common grid. This is useful to evaluate and use the resources.
- **Information column**: This column contains several icons:
  - ![folder icon]: indicates the dataset resource type. Series (dataset collection) are indicated by the folder icon (📁);
  - ![GEOSS icon]: the GEOSS icon shows which GEOSS category the given resource belongs to;
  - ![information icon]: the information icon provides resource information. By clicking this icon individual metadata for a resource are displayed, as shown in Figure 6.
- **Title column**: the title associated to the resource.

![Figure 6 – Resource metadata](image)
The above figures and comments describe the main features of GI-portal before the GeoViQua project. The next section describes the enhancements applied in the framework of GeoViQua, with the introduction of new concepts related to data quality specifications and quality-aware capabilities.

2.2 Enhancements to GI-portal

The DAB-Q Demo Portal is an evolution of GI-portal and introduces several elements for the quality and the feedback options.

The DAB-Q Demo portal is shown in Figure 7.

![DAB-Q Demo Portal](image)

Figure 7 – DAB-Q Demo Portal

Relevant changes to the portal regard the query constraints, the support to discover and access quality-aware services (e.g. CSW-Q, WMS-Q, SOS-Q), and the ability for users to provide their feedback on geospatial resources.

2.2.1 Improvements to the query constraints

It’s necessary to sketch the GeoViQua quality model before directly discussing modifications to query constraints in order to fully understand how the quality parameters have been chosen.

The GeoViQua quality model allows ISO 19115/19139 to be extended with richer quality information, reference to publications and documentation of discovered issues with a dataset.
Quality information can be stated both by data producers and by data users, actually resulting in two conceptually distinct data models, the Producer Quality Model (PQM) and the User Quality Model (UQM) or User Feedback model. For further details about the quality model see [GVQ-D6.1]. With the introduction of this quality model, it was logical to add some new query constraints in order to discover resources/datasets with associated quality aspects.

The DAB-Q Demo Portal has been improved to take into account the quality parameters. In particular, the “Query constraints selector” panel, on the left of the GUI, contains the additional “Show Quality and Feedback options” button for applying filters with quality-related constraints. The query constraints for quality and feedback options are shown in Figure 8.

The additional DAB-Q Demo Portal quality queryables operate either on the Producer or on the User Quality Model, as follows:

- **Producer Quality Model (PQM)**
  - Quality info count
  - Thematic accuracy result value
  - Domain consistency result value
  - Lineage process
  - Lineage source

- **User Quality Model (UQM)**
  - Rating
  - Report aspect
  - Main text
  - User domain
The chosen queryables are a small subset of the possible conceivable representations of data quality. This particular choice has been discussed with the GeoViQua partners, and represent a significant subset of the quality measures captured by the GeoViQua quality model. The next paragraphs detail the above queryables.

2.2.1.1 Support to query the PQM

The Producer Quality Model introduces indicators to record qualitative and quantitative quality information, and to identify resources (i.e. datasets) in order to relate metadata in hierarchical or other ways. The model extends ISO 19115, 19115-2 and 19157, adding means to report publications, discovered issues, reference datasets used for quality evaluation, traceability and statistical summaries of quantified uncertainty.

Quality indicators are conceptually classified into completeness, logical consistency, positional accuracy, thematic accuracy and temporal accuracy.

Table 1 summarizes the ISO standards quality indicators and its definition.

<table>
<thead>
<tr>
<th>Quality Class</th>
<th>Quality indicator</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positional accuracy</td>
<td>Absolute or external positional</td>
<td>Closeness of reported coordinate values to values accepted as or being true</td>
</tr>
<tr>
<td></td>
<td>accuracy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relative or internal accuracy</td>
<td>Closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true</td>
</tr>
<tr>
<td></td>
<td>Gridded data positional accuracy</td>
<td>Closeness of gridded data position values to values accepted as or being true</td>
</tr>
<tr>
<td>Completeness</td>
<td>Commission</td>
<td>Excess data present in a dataset</td>
</tr>
<tr>
<td></td>
<td>Omission</td>
<td>Data absent from a dataset</td>
</tr>
<tr>
<td></td>
<td>Conceptual consistency</td>
<td>Adherence to rules of the conceptual schema</td>
</tr>
<tr>
<td>Logical consistency</td>
<td>Domain consistency</td>
<td>Adherence of values to the value domains</td>
</tr>
<tr>
<td></td>
<td>Topological consistency</td>
<td>Correctness of the explicitly encoded topological characteristics of a dataset</td>
</tr>
</tbody>
</table>
The chosen PQM queryables are listed in Table 2.

**Table 2 – Producer Quality Model (PQM) queryables**

<table>
<thead>
<tr>
<th>Name</th>
<th>Queryable name</th>
<th>Quality parameter indicator</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality info count</td>
<td>QualityInfoCount</td>
<td>Completeness</td>
<td>Integer</td>
</tr>
<tr>
<td>Thematic accuracy</td>
<td>TCC.ResultValue</td>
<td>Thematic accuracy</td>
<td>Double</td>
</tr>
<tr>
<td>result value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain consistency</td>
<td>DC.ResultValue</td>
<td>Logical consistency</td>
<td>Double</td>
</tr>
<tr>
<td>result value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lineage Process</td>
<td>LI.Process</td>
<td>-</td>
<td>Integer</td>
</tr>
<tr>
<td>Lineage Source</td>
<td>LI.Source</td>
<td>-</td>
<td>String</td>
</tr>
</tbody>
</table>
The *Quality info count* queryable indicates the number of quality information associated to a dataset. The *Thematic accuracy result value* queryable represents the misclassification rate of the dataset. The *Domain consistency result value* queryable denotes the value domain conformance rate of the resource. Both are expressed as a real number (in double precision). The *Lineage Process* and the *Lineage Source* queryables are used for qualifying the lineage of the resource.

### 2.2.1.2 Support to query the UQM

The User Quality Model (UQM) describes the structure and attributes of comments, citations, discovered issues, ratings and reports of usage. It re-uses some ISO quality and metadata elements, and elements of the producer model, but is less strictly bound to existing ISO schemas.

The chosen UQM queryables are listed in Table 3.

<table>
<thead>
<tr>
<th>Name</th>
<th>Queryable name</th>
<th>Quality parameter indicator</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>Rating</td>
<td>User quality</td>
<td>Integer</td>
</tr>
<tr>
<td>Report aspect</td>
<td>ReportAspect</td>
<td>-</td>
<td>Codelist (usage, problem, alternative, fitness for purpose)</td>
</tr>
<tr>
<td>Main text</td>
<td>mainText</td>
<td>-</td>
<td>String</td>
</tr>
<tr>
<td>User domain</td>
<td>UserDomain</td>
<td>-</td>
<td>URN</td>
</tr>
</tbody>
</table>

The *Rating* queryable indicates the score/rating associated to a dataset. The possible value is an Integer between 1 and 5. The *Report aspect* queryable denotes the category of the report (possible values are: usage, problem, alternative, fitness for purpose). The *Main text* queryable is used for searching the specified string in the main textual elements of a feedback. The *User domain* queryable denotes the domain of interest of the users whose feedback is queried (e.g. Hydrology).

### 2.2.2 Improvements to quality-aware services

The DAB-Q Demo Portal allows brokering services implementing the quality-enabled interfaces defined by GeoViQua, like WMS-Q, SOS-Q. Figure 9 shows examples of such additional service instances connected to portal. In particular, the WMS-Q services provided by CREAF and UREAD, the SOS-Q service implemented by 52North, and the GeoViQua catalogue (WAF).
WMS-Q is a convention for encoding quality information in a Web Map Service; the only extension to the WMS 1.3.0 specification is minor (the modification to the MetadataURL element). The OGC Engineering Report “OGC® OWS 9 Data Quality and Web Mapping Engineering Report” (OGC 12-160r1) was formally approved on 18 January 2013.

SOS-Q service is a quality-aware extension of the OGC Sensor Observation Service 2.0 (SOS 2.0) specification. It is based on a mapping of SensorML-Q to the GeoViQua quality model.

The GeoViQua Catalogue (WAF) contains files of metadata records based on the GeoViQua Producer Quality Model (PQM). Duplicates of some of these records can also be found as examples in following folder: http://schemas.geoviqua.org/GVQ/4.0/example_documents

2.2.3 Improvements for feedback support

GeoViQua users may submit their rating, comments and other quality evaluations on datasets and other geospatial resources. This user-generated quality information is stored by the Feedback Catalogue.

The DAB-Q Demo Portal allows interacting with the Feedback Catalogue through its service interface. In particular, the users can post their feedback on the geospatial resources returned by query operation.

An example of query results is displayed in Figure 10.
Two icons implement feedback support in the DAB-Q Demo Portal:

- ![Q](image)
  - the icon represents the Quality metadata info. Only datasets with associated quality information exhibit this icon. Clicking on the Q-icon a GetRecordById operation is executed through the CSW-Q interface and the quality metadata for a resource is displayed.

- ![U](image)
  - the icon represents the User Feedback info. Clicking on this icon, a pop-up window allows the user to submit feedback on the dataset.

The User Feedback editor window is shown in Figure 11.
Users can insert the score/rating related to the resource (by marking 1 to 5 stars), a comment, their domain of interest (e.g. hydrology, seismology), and the category of feedback report (e.g. usage information, discovered problems). These aspects correspond to information elements of the User Quality Model (UQM).
3. GeoViQua GEO Portal Mirror

The GEO Portal is a website that provides convenient discovery to the full range of GEOSS data and information. To do this it dialogues with the DAB that exposes a metadata catalogue interface to the client. Operated by the European Space Agency (ESA) and the Food and Agriculture Organization of the United Nations (FAO), it provides a web-based interface for searching and discovering datasets, imagery, services and applications.

3.1 GEO Portal before GeoViQua

The operational GEO Portal that has been in place for the duration of the GeoViQua project is developed by ESA and is publicly available at www.geoportal.org. It mainly consists in a web interface to search the GCI catalogues and brokers. The current version of the start page that a user sees when accessing the site features a search using a “text only” control in a way similar to Google search. The search box automatically suggests search terms based on the entered characters. This page also contains editorial content and links to extra pages that will show a prepared list of search results for certain topics. We will call this first page the “entry page” (Figure 12).

![Figure 12 – Current GEO Portal entry page.](image)

After doing a search, e.g. by typing a word in the search box of the entry page and pressing “go”, a page with a statistics of the search results and the first block of ten results with a brief description of them is presented after a search or after a search refinement. We will call this page the “Search results/dataset listing page” (Figure 13). Users can browse the search results with common controls to flip through the search results “pages”. The brief search result description consists of a short text and iconic representation of the type of resource, such as “dataset”, “website” or “organisation”. A legend for these symbols can be opened by clicking on a link. For some results additional icons are displayed showing certain available resources for each item; for example a “chain” symbol if the result references an online resource. This is the first of two tabs of the search result page.
The second tab is called “Refine this Search”. Currently, it is possible to refine the search results by region, societal benefit area (SBA), keywords and dates. We will call this the “Refine this search page” (Figure 14).

One search result can be selected from the search results page for further inspection by clicking on the title of the search result item. This displays a new page that presents metadata about the selected result in three different forms: a tabular overview, a table of all metadata elements, and the raw metadata document. We will call this page the “Single result /dataset view page” (Figure 15).
Sometimes, it is also possible to view or access the data (if this is available for this layer). For example, if the layer can be seen as a WMS service (technically speaking if the dataset metadata has a WMS resource associated with it in the “distribution” section), a web map browser is shown. We will call this page the “Geo map browser” page (Figure 16).

3.2 GEO Portal improvements

One of the objectives of the GeoViQua project is to demonstrate that, with simple modifications in some pages, it is possible to include quality enhancements in the GEO Portal to make filtering and visualization more productive and attractive. To do that, Sapienza Consulting, as ESA subcontractor for the GeoViQua project, has provided a mirror copy of the GEO Portal that is being modified by the project members, and centralizes several visualization improvements. It is available online at http://scgeoviqua.sapienzaconsulting.com/.

Next, we are going to explain the suggested modifications for each of the pages described before except the “entry page” that shall remain unaltered. Since the live demo is available online under the link above and we try to incorporate as much feedback as possible the screenshots below might in some cases not show the final version of the developments with respect to user interface design.
3.2.1 Modifications in the Search Results / Dataset Listing page

The Search results page is the part of the portal where we see the highest potential of improvements by incorporating visual components for quality information. It has been modified in several ways.

3.2.1.1 Visual Metadata Summary

First thing that is needed is to present more information about each record that matches the search criteria must be presented in an appealing and quickly graspable way. Therefore we will add a small GEO label to each search result. For details on the GEO label see [GVQ-D6.2] and http://geolabel.info/.

The GEO label is a visual summary of available metadata and covers aspects such as metadata compliance, format, availability of reviews and user feedback. The label allows a visual comparison of the search results and is highly interactive, i.e. a user can access detailed information by hovering over the label or clicking on parts of it. The need to show a GEO label for each dataset is displayed in Figure 17.

![Figure 17 – Preliminary search results page (GEO Label option)](image)

3.2.1.2 Result Item Actions

Another important feature is the possibility to select more than one result with a check box and execute an operation based on the underlying metadata document(s). The current list of possible operations for one or multiple results starting from a search results listing page are described in this section.

The possibility to select more than one result item and the action dropdown button showing the possible actions available for more than one result simultaneously are shown in Figure 18.
Metadata Comparison

A screenshot of the metadata comparison showing a map and color-coded comparison of specific metadata elements of two example datasets is displayed in Figure 19.
Figure 19 – Metadata comparison

**Provenance Viewer**
The provenance view is a specialized visualisation of the lineage metadata that is available in a metadata document. It displays the lineage as a tree of process-steps and sources.

**Metadata Rating**
The “Rubric-Q” allows to assess the completeness of the metadata with a visual representation and a numerical indication/score.
3.2.2 Modifications in the Single Resource / Dataset View page

Besides the comparison of (meta-)data, several visual representations of a single dataset have been developed. These new methods are partly accessible from the search results page, but most prominently from the dataset view page. Currently this page does not have tabs but just some dynamic texts at the top of the page. The following new features shall therefore be integrated as new tabs or as extensions to the existing dynamic texts as shown in Figure 20.

Visual Metadata Summary

The page will contain the full GEO label in a larger size compared to the search results which facilitates the interactive features and the drilldown characteristics.

Quality-enabled Viewers

The page shall also show buttons for accessing WMS or WMS-Q in the new map viewer Greenland. The WMS-Q feature allows to directly showing a visualisation that is suitable to explore the quality of a raster dataset. A KML-Q button allows opening a specialized visualisation for vector datasets given in KML-Q format in a virtual globe viewer. Here we consider the possibility to use the Google Earth plug-in; as the “entry page” already does. For both of these adapted visualisations to work the metadata must contain the respective annotation.
User Feedback
User feedback is one of the core additions to GEOSS by GeoViQua. Apart from the GEO label, which displays the availability of user feedback, a specific sub-page for user-generated metadata will be added to the dataset view page. Here a user can browse other user’s comments and ratings of a dataset based on the information provided by the Feedback Server (http://twiki.geoviqua.org/twiki/bin/view/GeoViQuaIntranet/Feedbackserver) through the DAB-Q. A user can also click on a link to submit his own feedback. Additionally, new metadata visualizations are included such as: “Feedback Items” (that allows either to browse the user feedback items or add/submit new feedback).

Metadata Visualisation
The features “Metadata Rating” and “Provenance Viewer” described above are also accessible from the single resource page view. “Rubric Score” allows assessing the completeness of the metadata, and get a numerical indication/score for it and the provenance view presents the lineage as a tree of process-steps and sources.

3.2.3 Modifications in the Refine This Search page
Before GeoViQua modifications, Refine This Search was only able to add to the previous query a filter by area restriction, by SBA restriction, by keyword restriction and by date period. With GeoViQua modifications, search can be also constrained by a quality criterion. Quality criteria can be expressed as a quality measure threshold that will result in a request to the producer quality model or a user feedback model. Values of the threshold can be provided by the user as a number but the user will also be able to specify them in a more interactive way using a slider. For the user feedback model, a rating threshold can be requested by illuminating the right number of the set of 5 stars. Figure 21 shows the new characteristics in it, such as the refine by quality, by rating and the introduction of statistical plots of the search.

Additionally, the old search criteria controls will be complemented by statistical value representations and ranges to assist the user in providing the right values. This will be explained in detail in section 4.1.
3.2.4 Extensions of Web Map Browsing

To allow advanced visualisation of quality data, GeoViQua extended the Web Map Service standard with a set of metadata annotations that allow clients to know the semantics of layers. These semantic metadata can be used for an adapted visualisation of the geospatial data that suits the use case of understanding the quality of a dataset. This profile and extension of the WMS is called WMS-Q. Therefore it was added an additional more advanced map viewer to the Geoportal to handle the specific advantages of the WMS-Q. It is called Greenland⁴ and allows an ad.

Since Greenland is for advanced users we decided not to replace the existing map viewer but to integrate it for exploration of WMS-Q datasets. Therefore an additional button is added to WMS-Q: resources that, in the same manner as the regular WMS viewer button, is displayed on resource listings and resource detail pages. Adding this button depends on the provision of a set of metadata elements in the resource description.

After clicking this button the user is forwarded to a dedicated page (see Figure 22) that displays the resource the user was looking at.

⁴ https://wiki.52north.org/bin/view/Geostatistics/Greenland
Figure 22 – Greenland map analysis client in the Geoportal
4. Search assisted by statistical plots

Metadata statistical plots are a useful way to explore and search large geographical datasets. Due to the increasing amount of geographical data supplied for a great variety of producers and providers, the selection of data appropriate to the purposes of the users is becoming a complex task. For example, a search for the word “temperature” in www.geoportal.org returns 144235 results (Oct 14th, 2013). It is clear that mechanisms to refine the search are needed. Currently, the Geoportal offers refine by location, by SBA and by time that can be used for making the selection narrower, but even though, a selection of temperature in Europe, in the last 3 years and for understanding climate SBA returns 240 results that is practical but still too much.

The purpose of the Spatial Data Infrastructures (SDI) and the GEOSS Common Infrastructure is to facilitate the discovery and use of geospatial information through, among other things, a catalogue service. These initiatives offer access to geospatial data archives, but offer only basic levels of interactivity and user assistance (Göbel 2003). As concluded by Ahonen 2005 in her PhD work, metadata services in the context of spatial data infrastructures should provide visualization tools for users to gain better insight to geographic datasets that are available to shared use. However, the guidance for using these tools should be considered carefully (Ahonen 2005).

This deliverable proposes new refinements by quality indicators and, in this section, we propose the use of graphical or statistical plots to assist the user in better finding a narrow set of datasets.

This chapter offers first a review of some SDI catalogues providing great amounts of information and then analyzes the search tools provided and the levels of use of graphical visualizations to discriminate the data. Additionally, the current scientific literature is also reviewed and innovative graphics are also collected as possible alternatives to select the best result. The possibility of adopting this type of search refinements in the GEO Portal is discussed.

4.1 Including search by statistical plots in the GEO Portal

The current GEO Portal does not benefit from statistical information and graphical plots but other geospatial portals apply this approach. This section reviews some geospatial portal examples with statistical aids.

4.1.1 Spatial Data Infrastructure of Catalonia catalogue portal

One of the tools that the IDEC (Spatial Data Infrastructure of Catalonia) provides to its users is a metadata repository (a catalogue) that includes metadata from a great variety of suppliers, from enterprises to local and national administrations, including institutions and organizations from different sectors. The Infrastructure has a portal that allows, at a first level, searches by keywords, both inside the total metadata records as restricted by resource types (datasets, services, images, sensors and viewers). The keywords can also refer to a location (Where) and to a theme or type (What). A search by institution is provided by list navigation and by keyword.

Once these textual filters are applied, a results page showing the metadata matching the criteria selected in the first step is shown (Figure 23a. and b.). The image below displays the appearance of the results page following a simple keyword query, where options to continue filtering and browsing the catalogue are provided with some of their statistical summaries. Let's analyze a bit this interface and the possibilities offered.
Obs.1: The results interface is not defined to be viewed fully in a single screen, and therefore to view all the page the user has to scroll down the screen (Figure 23 a.). This performance is not too useful for being aware of what information is available in the catalogue. Users can be confused thinking that only a few results exist that fit their needs. Only when the bottom page is reached, a series of buttons allow browsing more results pages (Figure 23 b.).

![Figure 23 – a) Top of the results page with only the few first items visible. b) Bottom of the results page where a great amount of pages containing lists of results available are displayed.](image-url)

Statistical representation type 1: In the results interface, a left menu with new classified and available filter options is offered. The menu shows a filter criteria based on what type of data is being searched with the different possible values and the number of files available in each case. Thus, the number of records can be seen before applying the filter. This is a statistical feature that is very useful for the user since it guides him to formulate the right question and avoids that user ask for conditions that will not return results beforehand.

Dataset Filters: Sets, Types (grid, txt, tin, vector), Dataset Scales, Years, Series, and Providers.
**Services Filter:** Sets, Types (CSW, WCS, WFS, WMS, WMTS, WPS, TileCache, SOAP) and Providers.

**Image Filter:** Sets, Types (different sensors), Years and Providers.

**Sensors Filters:** Sets, Types (different radiances, wavebands,...) and Organization.

**Visors Filter:** Sets, Types and Providers.

When filtering by using the categories of the left menu, the list of datasets matching all the aggregated criteria is shown at the centre of the page. There is some visual information on each dataset (see the next figure): name of the dataset (orange arrow), snap-shot (blue arrow), availability (red square) and a link to a new server which has additional basic background information displayed together with the new dataset (green square).

All this information is shown as:

These features are a very preliminary use of the statistical and visual assisted search. Later in the text some improvements and more intuitive tools to display the information available and to facilitate the browsing task will be suggested.

**Obs. 2:** There is more information visible on the map view (yellow square in Figure 25), which defines what is being shown on the map: Datasets, Services, Images and Sensors. This information is also available in the top of the first page. Here, the geographical distribution of the available data is shown in a visual interface, allowing consulting and picking the desired items.
4.1.2 The INSPIRE Geoportal

Another example of metadata catalogue is the INSPIRE Geoportal: [http://inspire-geoportal.ec.europa.eu/discovery/](http://inspire-geoportal.ec.europa.eu/discovery/). The INSPIRE Geoportal provides the means to search for spatial datasets and spatial data services, and to view spatial data sets from the EU Member States within the framework of the INSPIRE Directive. In the next paragraphs we offer a review of the main features of this catalogue and its usability. Figure 26 shows the home page of the portal.

![Figure 26 – Home page of the INSPIRE Geoportal](image)

Initially, the catalogue offers the possibility to search by geographical location or by keyword. You can write a place in the left box or a theme in the right box.

Statistical representation type 2: The keyword cloud is another form of statistical representation of the different keywords that the user can select. The size of the word is a function of the number of
times that the word is used in different datasets. This is visually helping the user to make the right request. The bigger the word is, the less specific the filter will be and more results will be obtained, so it is better to choose the words with lower font size even if this can be a bit counterintuitive at the beginning. The words include both geographical names and thematic tags and show the number of items classified with each tag when leaving the mouse focus over them.

Obs. 1: In fact, the geographical search doesn’t work without the definition of a bounding box, and this is not explained in the page. Writing a name in the box only locates the site in the map, but do not apply a filter to the search criteria in the right panel. Once the site is located, you must draw a bounding box in order to restrict the results to that place (Figure 27).

![Figure 27 – Results of the application of a geographical filter by drawing a bounding box.](image)

Obs. 2: Both the geographical search and the text search have a pop-up menu showing the possibilities existing in the database. This is a visual tool very useful to help finding the right words when describing sites or themes, while it gives an idea of what can be found. Both search boxes are also multilingual, i.e. you can search in any language of the European Union. This form of assisting the user is also present in the GEOSS GEO Portal. It guides the user into using more specific words by offering more concrete concepts that contains the requested word (Figure 28).

38
Obs. 3: As can be observed in Figure 28, the list of words suggested by the pop-up menu doesn’t include the number of items existing in each category, as is explained in the help balloon that appears when placing the mouse over the box (red arrow). The use of this statistical feature would contribute to facilitate the adoption of graphical search tools.

Obs. 4: The search can be also filtered from this page by the type of data, including datasets, series, services, layers or download services directly plus combinations of types.

Obs. 5: The advanced search permits to refine the search by applying filters to the origin, metadata language, spatial data theme, topic categories or service type (Figure 29). These filters also work like a pop-up menu and each category shows the number of results available. This is the same feature shown in the IDEC portal. See Statistical representation type 1.
Obs. 6: Once the search is performed, there is the possibility of sorting the results by relevance, resource title, origin, date of creation, date of last revision, resource type or service type (Figure 30). The current Geoportal is no capable of sorting the results by relevance.

4.2 Proposals on the use of statistical plots

Both the IDEC and the INSPIRE geoportals use the classification of the data into categories and some statistical features like presenting the total sum of the datasets related to the different values of some categories or search criteria, normally in parenthesis. This type of feature involves the labelling of the data and its classification and then allows for the statistical analysis. This basic information and some more complex one can be shown in a more intuitive way to help find the adequate data for the user, for example by the use of graphic plots or drawings, as metadata is necessary in both textual and visual forms, since the use of multiple representations is a characteristic of geospatial metadata that is typical in exploratory environments (Ahonen 2005). In this section some proposals on the use of this kind of statistical plots in the reviewed geoportals are presented.

Some additions that could be incorporated to this and to other geoportals and that are considered in some papers include searches based on the diagram representation of the categorized information that helps in the comparison and the exploration of data during the search. The use of visual based approaches for metadata analysis is important to facilitate the users to locate appropriate data (Albertoni 2004) and have been well evaluated in users tests (Ahonen 2005). These approaches combine automatic visualization techniques with graphic interaction tools to create a data exploration system, although instruments able to manage this large set of metadata and their multidimensionality are needed.
The intention is to increase the use of the visual analysis of metadata to help users to choose between different sets. The most basic and used features, implemented in the reviewed geoportals (both IDEC and INSPIRE) are the representation of the attributes by categorical values or full text values and the use of a bounding box or a point in a map to express spatial extent attributes. But these features can be improved by the adoption of graphical representations of the combined categories to dynamically construct the searches according to the exploration of the information that let to go successively refining the list of adequate results in a multi-step information retrieval process: query formulation, search result presentation, data analysis/comparison and query modification (Göbel 2003). Shneiderman presents the process of seeking visual information as "overview first, zoom and filter, then details on demand" (Shneiderman 1998).

Although there are a lot of different proposals, including 3D simulations (Göbel 2003), the most used graphical features to present this type of information are based in a multidimensional representation of the values of multiple attributes or search criteria in a plot or in a visual way: the parallel coordinates plots, the scatter plot matrix, the star plots and the Chernoff faces (Ahonen 2005).

### 4.2.1 Parallel coordinates plot

The parallel coordinates plot uses different vertical axes to rank the values of each attribute in one axis and their display together in parallel in order to make easy the comparison between the attributes and facilitate the choice of the right set of data for the purposes of the user (Figure 32). In this case, it is important that users can control the way in which axes are displayed because it affects the nature of the plot and possibly its interpretation.
4.2.2 Scatter plots

The scatter plot matrix uses a spatial representation of the different criteria by pairs, in one single matrix (Figure 33). To keep a matrix of scatter plots manageable the number of variables cannot be high. This kind of representation did not gain much acceptance in a test conducted by Ahonen (2005).

4.2.3 Star plots

The star plots are similar to the parallel coordinates plots, but in this case the axes are centred in the same point and extending in all directions (Figure 34). In these axes the values of the different attributes are represented, resulting in a kind of radial plot with different volumes according to the datasets. These volumes take the form of stars, therefore the name. Although in this case the ordering of the axes influences the shape of the star, this is usually not under user’s control, whereas user-control of parallel coordinate axes is a rule. It is also recommended not to use more than three orientations (Ahonen 2005).
4.2.4 Chernoff faces

In the Chernoff faces (Figure 35), each attribute is represented by a piece or portion of a face drawing and the shape and size of the piece is related to the value of the attribute. This type of graphic is very sensitive to the number of criteria represented and its size and shape because of the limited perception of the combination of attributes. For example, it is difficult to distinguish the shape of the eyes if their size is small. That is why it is recommended not to exceed three or five variables in the representation and also allow users to set the type of variables and their corresponding pieces in the drawing (De Soete 1986).

4.3 Factors to consider when implementing statistical plots

Even though the displayed usefulness of the visualization techniques in the exploration of metadata, some questions have to be considered, for example the problem of how to deal with missing values in the multivariate data representations. This connects with the idea of the GEO label that allows us to determine the completeness of our metadata before applying statistical plots. These representations are also more useful when implemented as multiple linked views (i.e. when the selection of objects in one view causes the same objects in the other views to be highlighted). In this way, relationships not directly obvious from any of the single views may become apparent, while multiple dimensions of a dataset can be distributed into separate but linked displays and be viewed simultaneously. And even though alternative views can be revealing, they may as well appear too complex for proper understanding to users not familiar with these non-conventional views (Ahonen 2005).

In this sense, it seems necessary to design a kind of representation technique with interactivity not only at the query level but also at the construction level. User must be able to configure the display
of the graphical information in order to make it understandable for him. For example, there must be the possibility of setting the parameters to compare and the value or importance the user gives to each one. It means the definition of the axes and attributes and their sequence or orientation in the case of a parallel coordinates plot (e.g. making that all criteria that are good for the user are represented at the top of the graphic), a scatter plot or a star plot (e.g. making that all criteria that are good for the user contributes to increase the area of the plot), or the identification of the different graphical symbols with the values of the attributes in the case of a Chernoff face or other iconic display. In other words, to get the maximum advantage of the visual techniques to help in searching geographical datasets, the customization of the technique by the user is desirable.

This kind of statistical plots is designed to be used with numerical values. Given that many of the attributes in metadata records are textual, the values of these variables need to be sorted and transformed or associated to a numerical value in order to be represented. For this reason, it is also useful to let the users establish these correlations according to their preferences, as the graphical appearance of the plot will become easily understood. In Figure 32 and Figure 34, the categorical values have been associated to ordinal values according to the desired requirements for the data and so the high values are shown in the top values in the parallel coordinates plot and the external values in the star plot, resulting in a comprehensive interpretation.

4.4 The GEO label as a statistical plot for metadata completeness

As mentioned above, one of the risks of the visualization techniques is that they are meaningless if no data exist. In a large dataset, the missing of single values it is not unusual. Some ways to face this peculiarity include indicating the missing values in the displays or replace them by plausible values (Ahonen 2005) but this is often not possible in metadata record. In any case, dealing with missing values or made up values doesn’t contribute to get an idea of the best datasets for the user, since we don’t know some of the information for these datasets.

In GeoViQua context, the idea of the GEO label has been developed, that could help the user in the proper comparison of datasets. The GEO label (http://geolabel.info) is a graphical representation of the availability of quality information for a given dataset and, for extension, of the missing values for this information. The GEO label consists in a circular graphic divided into 8 sections, each one related to an informational facet: producer profile, lineage information, producer comments, compliance with standards, quality information, user feedback, expert reviews and citations information. Each informational facet can represent one of three availability states: ‘available’, ‘not available’, and ‘available only at a higher level’ (indicates that information is not immediately available for the dataset, but is available for a parent dataset). These three information availability states are expressed through varying the appearance of the facet icons (see the Figure 36).
The use of the GEO label in a dataset selection process can help in the overview and filtering of the information in some cases. As concluded by Ahonen (Ahonen 2005), the use of multiple representations is a typical characteristic in exploratory environments. According to this situation, the GEO label could be used as a first exploratory level, filtering the completeness of the metadata to be browsed. Then the statistical plots would be applied only to these datasets with a sufficient level of completeness in the metadata values, in order to avoid missing values in the plots. Another use of the GEO label would be in the final step of the selection, helping in discriminate between a group of datasets that are adequate to the purposes of the user for other reasons, or even in the same statistical plots as another attribute of the metadata: their completeness.

**4.5 Example of implementation in the GEO Portal Mirror GUI**

In the context of the GeoViQua project, the use and application of this kind of search tools has been implemented in the GEO Portal Mirror GUI, in order to test their performance and effectiveness in assisting the search for datasets. For this preliminary implementation, a group of similar datasets has been selected and used to apply the metadata comparison and the statistical plots to differentiate the datasets and their attributes so that to choose between them could be done in an easier and assisted way.

The datasets to compare can be obtained from a previous search in the GEO Portal using the current search tools available (search by keyword, bounding box, date or SBA). Then, in order to refine the search, the new options implemented in the GeoViQua project will be used. The first visual tool that allows for evaluating the data is the GEO label. It appears in the list of results giving visual information about the completeness of the metadata for each record. Figure 17 of this deliverable illustrates this new feature developed in the GeoViQua project. This visual tool helps to be aware of the amount of the information available about the data before proceeding to compare the metadata. In other words, if the GEO label appears in white, there is no metadata to compare with other data record or to learn more about the data. Keeping this in mind, it is appropriate to select the best records of the list in terms of metadata completeness in order to enable the metadata comparison to be effective.
Once the selection is done, the possibility of performing a metadata comparison is offered in the submenu as shown in Figure 37.

![Figure 37](image)

Figure 37.View of the submenu implemented in the GEO Portal Mirror GUI for the metadata comparison.

As explained in a previous section of this deliverable, the metadata comparison gives a visual list of the data attributes with each dataset organized in columns, so that the information for each record is available and side by side comparable between the selected items. This feature also offers a location map of the bounding box covered by the dataset, based on the extent information registered in the metadata, which contributes to increase the visual appearance and support of the tool. The metadata comparison has been described in the section 3.2.1.2 and showed in the Figure 19.

Finally, to complete the availability of the graphical search, the possibility of generating statistical plots of the selected records has been added to the GEO Portal Mirror GUI by integrating it in the metadata comparison tool. It means that it acts automatically over the same group of selected records to compare and that the attributes plotted are predefined. In the future a interactive way of selecting attributes is foreseen. The plot display is added to the metadata comparison interface, as shown in Figure 38.
Figure 38. Implementation of the statistical plots in the metadata comparison interface of the GEO Portal Mirror GUI. View of a star plot at the left side and a parallel coordinates plot at the right side.

As seen in the Figure 38, there are two types of plots available for the comparison: the star plot and the parallel coordinates plot (see section 4.2 for more information about the different types of plots). The attributes to compare for instance includes the following: extent, absolute external positional accuracy, date, metadata date stamp and maintenance frequency. All these attributes have the characteristic of being numeric or in any way sortable. The implementation converts all the values into a 0-100 rank depending on the goodness of the value. For example, a global extent (like a worldwide extent) is given a 100 score while local extents are given smaller scores in order to indicate width. The external positional accuracy is valued according to its meaning: the best positional accuracy is given 100 and the worse are given smaller values between 0 and 100. In the case of the dates, the most recent dataset is consider better and so scored 100 and the oldest dataset is given 0. The maintenance frequency is scored 0 to the unknown value and 100 to the continual value, with a scale of sorted frequencies. These assessments result in a big surface in the case of the star plot and an upper line in the case of the parallel coordinates plot for the best datasets, making the comparison easier to understand.

The example of the Figure 38 is obtained from a group of datasets of the GEOSS catalogue in the GEO Portal. All of them are related to oceans, but each one has its own characteristics. To differentiate them, it becomes very useful to analyze their attributes through the metadata comparison and the statistical plots, given that the list of results offered up to now in the GEO Portal does not allow having an overview of these characteristics, unless individually. Thus, it is easy to see from the maps that, in the data group, there are one local dataset (from Cuba and Puerto Rico), one middle-extent dataset and three global datasets. Also we know from the text that we have data related to aerosol optical thickness, to cetacean surveys, to coral bleaching monitoring, sea surface temperature or ozone capture. Three of these datasets are included in the climatology/meteorology/atmosphere topic category, one in the oceans topic and the last in the biota topic. We also can see the publication date and the contact information of each dataset or the quality information of the data.
FP7 Project Nr: 265178  
Acronym: GeoViQua  
Project title: QUALity aware Visualisation for the Global Earth Observation system of systems  
Theme: ENV.2010.4.1.2-2  
Theme title: Integrating new data visualisation approaches of earth Systems into GEOSS development
From the statistical plots, we can understand that we have a variety of dataset extensions and that all the data have the same maintenance frequency “as needed”, but then we have some recent datasets and some old ones, and especially one with a low positional accuracy. Assuming that all datasets were providing information about the same thing, and unless we were interested in a local dataset, the options 1, 3 and 5 would be the best choices for us, giving that the dataset 2 has a local extent and is the oldest (although also has the better positional accuracy) and the dataset 4 is good in terms of the extent and the date but has the worst positional accuracy.

As seen in the literature discussed in the previous subsections, the use of multiple graphical options when searching for data helps to improve the knowledge about the datasets and therefore the possibilities for a better choice. Accordingly, the implementations proposed in this deliverable include a variety of graphical and visual tools offering different points of view of the datasets: the GEO label for the completeness of the metadata, the metadata comparison for viewing the main information of each dataset in a glance and the statistical plots to evaluate the better dataset of all. A good practice would be to search first by keyword or SBA, setting the date and the bounding box, review the GEO labels to choose the desirable metadata and then ask for the metadata comparison (for being aware of the qualitative information) and finally examine the statistical plots (for helping in the choice of the dataset with better score).
5. Conclusions

5.1 Recommendations for GCI integration

5.1.1 Recommendations about the inclusion of GeoViQua broker
We recommend to discuss with the IIB (Infrastructure Implementation Board) and the DAB manager to incorporate the GeoViQua broker, the Feedback catalogue and the CSW-Q component in the GCI, as a contribution of GeoViQua project, for supporting users in quality enabled filtering and querying.

5.1.2 Recommendations about inclusion of statistical plots
Once a simple search has been done a filter criteria in the advanced search interface is possible. The interface does not present any clues on the possible filtering values on the filter options. It could be helpful that for each criterion, the different possible values are listed and the number of records foreseen as a result of applying this filter value can be seen even before formulating the filter. This statistical feature is very useful for the user since it guides him to formulate the right question and avoids that user ask for conditions that will not return results beforehand.

The GEO Portal offers a list of keywords to search but no clue about how many results a keyword can get. The keyword clouds are another form of statistical representation of the different keywords that the user can select. The size of the word can be a function of the number of times that the word is used in different datasets. This is visually helping the user to make the right request. The bigger the word is, the less specific the filter will be and more results will be obtained, so it is better to choose the words with lower font size even if this can be a bit counterintuitive at the beginning. The words can show the number of items classified with each tag when leaving the mouse focus on them.

Metadata comparison tools are very useful to pick the right dataset out of a few possible candidates. A graphical multidimensional representation of the values of multiple metadata elements of search criteria in a plot or in a visual way is convenient. Examples of those are: the parallel coordinates plots, the scatter plot matrix, the star plots and the Chernoff faces. Both parallel coordinate plots and the star plots are recognized and understood better by the users but require the right interfaces to choose the comparison criteria and a way to deal with NODATA values. The integration into a textual metadata comparison tool might help with these aspects.

5.1.3 Recommendations about inclusion of the user feedback
The world wide web’s success story is a story about a web of contributors rather than a web of consumers. Therefore it is only natural to investigate collaborative efforts for quality assessment in the GCI as well. GeoViQua made considerable progress of integrating user feedback into the Geoportal with a powerful and cleanly specified service-oriented architecture for storage, management, and retrieval of user-generated metadata. We recommend to make this integration publically or semi-publically available to build up a collection of user feedback items, i.e. actual text comments, expert reviews and ratings, to facilitate further research in this area.

5.1.4 Recommendation about inclusion of GEO label
The GEO label is a well-designed visual summary of geospatial metadata from various sources. The complexity of deciding about fitness-for-purpose based on standardized metadata, expert reviews
and user comments is so high that we see the integration of the GEO label as a crucial effort to increase the usability of the GEO portal.

5.2 Future improvements

5.2.1 Query by label in the DAB-Q Demo Portal

The search interface may be improved with the integration of the GEO label in order to specify additional (optional) query constraints.

Figure 37 is a mock-up with the GEO label icon among the quality and feedback options.

![Figure 37 – Search interface with disabled GEO label filter](image)

As GEO label constraints are optional, the GUI control for GEO label queries can be toggled active/inactive. Clicking on the “Enable GEO label queryables” button activates the GEO label queryables (Figure 38).
When active, clicking over the GEO label facets will filter the results according to the highlighted constraints. For example, users could ask for datasets with “Producer profile”, “Producer comments”, “Compliance with standards”, “Quality information” and “Citations information” fields, as shown in Figure 39.
5.2.2 Customizable statistical plots

Another feature discussed above is about the need for customization of the statistical plots in order to make them actually useful for the interpretation of the users. The implemented graphics were intended to serve as a viability and demonstration test. It is obvious that the users should be able to decide which attributes to compare in the plot and what score to give to each value so the plot is reflecting exactly their wishes. Interactive customization of the statistical plots proposed is foreseen in future designs of the GEO Portal.

Figure 39 – DAB-Q Demo Portal with selected GEO label facets
6. References

[Ahonen-Rainio 2005]

[Ahonen-Rainio & Kraak 2005]

[Albertoni 2004]

[Albertoni 2005]

[Cox 2011]

[De Soete 1986]

[Göbel 2003]

[GVQ-DOW 2010]

[GVQ-D6.1]
GeoViQua Consortium, Deliverable D6.1 “Best practice document for quality encodings”.

[GVQ-D6.2]
GeoViQua Consortium, Deliverable D6.2 “GEO Label description final document”.

[ISO/IS 19119:2003]

[ISO/IS 19115:2003(E)]
[OGC, 2003]

[Portele 2007]

[RM-ODP 2009]

[Schut 2007]

[Shneiderman 1998]